

EXHIBIT D

Part 9 of 14

The intent of the Archambeault and Deis evaluation was strictly to compare the public versus the private operation of the facilities. If there is an innovation to be found, it might be the staffing patterns used by Corrections Corporation of America. The public facility used predominantly white (81.5 percent) males (76 percent). The CCA-operated prison used many more female staff (42 percent) and many more minority staff (50.3 percent). The WCC-operated facility was somewhere in between, using mostly white male staff (63 percent). Since staffing is one of the most important components of correctional service, the use of more women and minorities might be considered an innovation if it could be shown to be related to the successful management of the institution. If Archambeault and Deis did point to an innovation it would have been the management philosophy of the CCA-operated facility. They considered CCA's approach to be the least authoritative, involving staff in the organizational decision making and giving "employees a vested interest in the overall success of the prison organization (p. 66)."

It was difficult to conclude how or whether the Louisiana public correctional system had changed in any way in response to the privatization of its two facilities. The Secretary of Public Safety and Corrections did introduce a reporting system to monitor all of the institutions; however, this system might have been implemented in the absence of privatization.

New Mexico Evaluation

The discussion of the New Mexico evaluation is taken primarily from the report by Harer, Karacki, and Gaes (1995).

Charles Logan has published, "Well kept: Comparing quality of confinement in private and public prisons," as both a National Institute of Justice monograph and a condensed version in the *Journal of Criminal Law and Criminology* (Vol. 83, pp 577-613, 1992). Logan's comparison of private and public institutions is based on contrasting the operations of two women's prison in New Mexico and a Federal institution. Women prisoners in New Mexico were originally housed in a state facility which operated primarily as a diagnostic and orientation facility for men. To meet the needs of the female inmates, a private facility was built and operated by the Corrections Corporation of America (CCA) under contract to the state. Logan later collected a limited amount of data from a Federal women's facility to do a comparison of all three kinds of operations.

In both his monograph and paper, Logan first outlined a theoretical model, the Confinement model, which he used to form the basis for comparisons among the three institutions. He then proceeded to develop institution performance indicators (IPT's) which represented the various dimensions of his Confinement model. Logan's Confinement model is succinctly represented in his own words:

The mission of a prison is to keep prisoners--to keep them in, keep them safe, keep them in line, keep them healthy, and keep them busy--and to do it with fairness, without undue suffering and as efficiently as possible. (Logan 1992: 580)

Thus, the dimensions which Logan used to compare institutions were security, safety, order, care, activity, justice, conditions, and management. We have little quarrel with Logan's conceptual framework, although we believe there are critical goals not adequately addressed by this model. In

Logan's approach to corrections, inmate programs are merely activities to "keep them busy." But, as the mission statements of most modern correctional systems imply, rather than merely "to keep them busy," prison programs are intended to provide inmates with a positive influence in their lives, afford them the opportunity to improve their skills, and provide a socializing agent so that they come to accept the moral and legal norms of society.

One of the fundamental problems with Logan's study is that with only three institutions to compare, he was limited in the kinds of analyses that could reasonably be used. Furthermore, he was really comparing the management of essentially the same New Mexico female inmate population over time by two different management groups (public versus private) housed in two different facilities. Almost as an afterthought, he contrasted this before/after longitudinal assessment to a population composed of imprisoned federal women. The appendices to Logan's NIJ monograph also include reports, written by two consultants, which described the state and private institutions and the events culminating in the transfer of female inmates from the New Mexico State prison to the privately run institution.

Logan used a large number of performance indicators which were quantitatively combined; however, the large number of measures used does not compensate for the very small sample under analysis.

Other problems with this study, however, arise from the ways in which Logan analyzed and interpreted the institution performance indicators that he used to compare the operations of the three facilities.

The Probable Bias in Institution Performance Indicators Based on Staff Perceptions

Almost all of the differences which favor the private facility over the two public facilities in the Logan study were based on staff perceptions as measured by the Prison Social Climate Survey (Saylor 1984) administered to staff at the three institutions. There are many reasons why staff surveys could have biased the results in favor of the private facility.

Clearly, staff at the private facility would be keenly aware that the success of their employment could depend, in part, on the responses they provided to the researcher. This most obvious, and potentially fatal flaw of the research, was discussed only superficially and then dismissed in the Logan report.

A second biasing influence was the fact that many of the staff selected for the private facility had worked in the public facility previously. It is likely that they would want to justify their decision to leave the state government to work in the private facility by favorably responding to the survey questions.

A third biasing influence in the measurement of these perceptions was the fact that the private facility was brand new, with many new staff as well as some experienced former state corrections staff. The facility built by the private corrections company was considered to be very well designed and staff were excited by working in this new environment. Yet, one wonders how their attitudes would have changed over time as the challenge of working in a new and exciting environment gives way to the daily routine of operating a correctional facility. We were so curious about the possible change in attitudes of staff over time that we asked the Corrections Department of New Mexico if we could re-administer the Prison Social Climate Survey at the private facility two years after the private operation began. However, management at the private facility would not allow this.

Another important biasing influence on staff perceptions was the low response rate among state and Federal employees. The private facility response rate was 72 percent while the two public facility response rates were less than 50 percent. Data from the Federal facility were collected in the very first year the Prison Social Climate Surveys were administered for the entire Bureau of Prisons. In subsequent administrations, the Bureau of Prisons has achieved no lower than a 72 percent response rate and as high as an 88 percent response rate from 1989 through 1994. Low response rates could be construed as indicating that only staff who had negative perceptions of the institutions were motivated to complete the Social Climate survey at the two public institutions.

The last important biasing influence in Logan's comparison across the three institutions and, perhaps the most damaging, is that Logan did not have a full complement of measures for each institution for computing an overall index of "quality." In particular, approximately 30 percent of the measures that were used in the comparison were missing for the Federal facility.

To exemplify this problem, suppose we are asked to rank order three employees on their overall performance based upon 10 performance indicators. For two employees, data are available on all ten measures; however, for the third employee, we only have information on 7 measures. Knowing this, should we still try to rank order all three employees on the overall performance index? Obviously not, yet this is exactly what Logan proceeded to do.

Interpreting Institution Performance Indicators

Another problem with the Logan study was what we believe to be the often questionable way in which Logan interpreted the institution performance indicators (IPI's). In all, Logan used 333 IPI's, most of which were based on staff perceptions measured by the Prison Social Climate Survey. For each of the 333 IPI's, Logan assigned a ++, +, =, -, or -- value to indicate whether an institution was much better, better, equal, or worse, or much worse than the comparison institution(s). It is noteworthy that a set of indicators and measures of correctional effectiveness derived from the professional judgment of multiple correctional experts and practitioners was available to Logan, namely the American Correctional Association's (ACA) standards. Logan chose to ignore the ACA measurement scheme, preferring his own instead. Logan himself admitted that "Interpreting each measurement item was often difficult." He also pointed out that many items could quite legitimately be interpreted and scored in many different ways. Yet despite his initial cautious remarks, he proceeded to make specific judgments, a priori, about whether a measurement item was positive or negative without consulting correctional experts. We found some of his judgments somewhat naive. A few examples should demonstrate this. We will focus here on only the relatively objective measures he culled from institution records for the institution security dimension.

In his section on security, Logan examined a number of objective indicators related to inmate contraband, drug use, misconduct, staffing patterns, and furloughs. Other than his inclusion of furloughs, we think these are valid issues related to institution security. Unfortunately, the way these measures were interpreted by Logan makes little or no sense from any sound correctional management point of view.

Logan compared the state and privately run prisons by looking at the rate shakedowns (i.e., contraband searches) were occurring and the proportion of shakedowns in which contraband was found. He gave the state facility credit for performing shakedowns more often; however, he gave the privately run facility more credit for finding contraband a lower proportion of time. This latter

judgment makes little sense. It may very well be that the state officials were doing a better job of finding contraband, were less tolerant of contraband, or both. These explanations would have led to an opposite conclusion from Logan's. Whereas Logan saw merit in finding less contraband, it is also reasonable to assume that finding more contraband is a reflection of thorough and well-trained security staff..

A second example involved the privately run facility, as well as the state and federal facility. Logan found that among inmates suspected of using drugs, fewer tested positive for opiates in the Federal facility than in either the state or privately run institution. Yet, Logan gave the Federal institution a demerit for conducting fewer tests among inmates suspected of drug use, while giving the privately run institution the highest rating for conducting the most urinalysis tests. The important issue here is whether inmates were using drugs in the institution. Conducting more tests might be construed as wasting money. How could the Federal institution be given credit for having the lowest opiate usage rate and simultaneously be judged inadequate in its drug testing policy while achieving a better result?

Another objective dimension reported by Logan was the rate at which inmates committed significant misconduct incidents during the 6-month period of the study. This is an example where Logan used his judgment, rather than conduct a statistical test. He found that the privately run prison which had a 0 rate of significant incidents was judged better than the federal prison which had a rate of .01 significant incidents in a 6-month period. These rates were not statistically different from each other, yet in Logan's judgment, the privately run facility performed better on this measurement item. Since there is so much reporting discretion in these kinds of incidents, we are not sure it is a fair assessment to conclude that the privately operated facility had out-performed the publicly operated facility when the rates for both facilities were so low.

In his analysis of institution security, Logan evaluated furlough rates as well. He claimed that furloughs indicated the extent to which prison administrators exposed the community to dangerous criminals. He reasoned that the higher the exposure (i.e., more furloughs), the lower the rating for the institution. Many correctional administrators look favorably at furloughs as a way of easing an inmate's transition back into society. Most correctional systems use furloughs on a very limited basis for inmates that are already close to their release date and who are thoroughly screened to minimize the risk to a community. A study conducted by Harer (1994) demonstrated that prison furloughs are one of the best predictors of an inmate's post-release success, when controlling for other risk predictors. Inmates who receive furloughs for the purpose of establishing community ties before release are less likely to be re-arrested or have their supervision revoked for the first three years after their release.

The final objective measure Logan used in the security section of his paper was the custody staffing levels of the different institutions. The Federal prison received 2 demerits (--), compared to the state and privately run institutions. The private, state, and Federal prisons had inmate-to-staff ratios of 3.1, 2.3, and 8.1 to one, respectively. How is it that the Federal institution, which had the lowest urinalysis rate, and an extremely low incident rate, could be judged unfavorably because it used fewer custody staff, by far, than the other two institutions, yet still managed to operate a safe and humane environment? A related problem with this interpretation is Logan's failure to understand the management context. The Bureau of Prisons considers all of its staff to be correctional officers first. Thus, teachers and vocational training instructors have both responsibilities as instructors and custodial staff. Under this model, if safety and security is still maintained, why would you debit an institution? The real issue here is whether a performance indicator is an outcome or a process measure. In the case of security, the primary outcomes are those that measure whether inmates and

staff are at risk to be assaulted or harmed. Other measures, such as staff to inmate ratios, shakedowns, and even furloughs are process measures that should be regarded as related to outcomes, but not as outcomes themselves.

The bottom line on the comparison between the private and the Federal facilities was that on objective dimensions, the Federal facility seemed to perform better. Since we regard the perceptual measures as likely biased, comparisons based on these measures strike us as highly questionable.

Although Logan has done a remarkable job conceptualizing the dimensions of confinement, we strongly believe that his lack of institutional experience limited his ability to compare the correctional institutions on specific performance indicators, especially since he relied on his own judgment about those performance indicators. His lack of a fundamental understanding about how prisons operate and how they are managed limits the usefulness of "Well Kept."

Putting the Public-Private Comparison in Perspective

In considering the history that led to the transfer of all female inmates from WNMAC, which served as the publicly-operated facility in this analysis, to a new CCA facility, which stood as the privately-operated facility, our impression is that we would have been surprised had the new facility not been an perceived as an improvement over the past by those staff who were surveyed for the purpose of this study.

First of all, as DiIulio commented in an appended section of the report, before the CCA facility opened, "New Mexico's women prisoners could be described as correctional 'orphans' who were housed in 'a make-shift' wing of Western New Mexico Correctional Facility, a large high-security institution for males." He further added that, "Before that, the women prisoners experienced frequent moves between different facilities, none of which was equipped to meet the needs of female inmates." DiIulio's comments seem to indicate that the conditions of confinement for the female offenders were not particularly favorable.

Moreover, when it is considered that the State correctional system was under court order to improve conditions (WNMAC in June 1988 was found to be in non-compliance in over one-third (35.7 percent) of the 42 compliance provisions audited), it certainly appears that problems existed in the operation of WNMAC. Our sense is that while State of New Mexico prison authorities attempted to provide for the female offender population at WNMAC, there were serious limits as to how successful this effort could be. This is suggested in the comments of Charles W. Thomas, a correctional consultant, who, in his assessment of WNMAC in October 1988, was very positive about institutional staff but was highly critical about institutional design features and the security problems these design flaws created. Indeed, our assumption is that the contract with CCA to operate a new female facility was intended, in part at least, to overcome problems which existed at WNMAC.

The point is that WNMAC was not just any State-operated facility for female offenders, but was instead a facility with major inadequacies, and CCA, far from being just another privately-operated female facility, was intended to replace WNMAC and presumably in the process to overcome the inadequacies of WNMAC. Under these circumstances, could CCA be anything but better?

Other Methodological Problems

In this last section, we list other methodological problems with the Logan study.

- ┆ Logan gave equal weight to each of his quality dimensions and sub-dimensions. Thus, a rating on security was considered just as important as a rating on activity. Most correctional experts consider security and safety to be the primary objectives of sound prison practices and would not rate all dimensions equivalently.
- ┆ Logan gave equal weight to each of his empirical measures, for example, the serious assault rate had no more weight as an index of "OVERALL" prison quality than how often inmates used the recreational facilities.
- ┆ He made no attempt to show if, and by how much, the subjective measures predicted, or were associated with, the objective measures.
- ┆ He ignored the magnitude of differences in item scores between institutions, forcing quantitative (interval scale) differences between institutions into a tripartite (+, -, and =) scale, while often making conceptually questionable decisions about whether high or low on the item score means "+". He then combined and re-quantified the tripartite difference measures using a conceptual method of scoring similar to ice hockey. We question the adequacy of this measurement process. A quantitative method could have been used.
- ┆ Although he purported to be looking only at three all-female institutions, Logan, without acknowledging it, introduced survey response bias by including responses by staff at the New Mexico State prison who worked with male inmates when making comparisons of staff Climate survey responses across institutions.
- ┆ Inmate survey data showed that the public New Mexico State facility outperformed the private facility in every dimension except "activity" (GAO report p. 28). This was contrary to the results based on staff perceptions. Why was there this discrepancy between inmate and staff perceptions? Are we to discount inmate perceptions based upon some underlying vested interest? Or perhaps the greater vested interest lies with the staff who had a financial stake in the success of the evaluation.

Although we find Logan's conceptualization of a theoretical framework from which to compare institutions appealing, for the reasons given above his attempt to evaluate the relative merits of public and private facilities falls far short of a rigorous or conclusive analysis. There was no indication in the report whether the State of New Mexico had changed its policy or procedures in response to the privatization of its female population. Furthermore, there is no indication in the report that there had been any innovation on the part of CCA in its management of the women prisoners.

Florida Evaluation

A recent study by Lanza-Kaduce and Parker (1998) compared the recidivism rates of inmates released from prisons operated by the Florida Department of Corrections to those inmates released from a

prison operated by the Corrections Corporation of America and a prison operated by Wackenhut Corrections Corporation. The at-risk release period for the offenders was 12 months and recidivism was defined either as a rearrest, a new offense, a new commitment, a technical violation, or a summary measure based on the first four indicators. Inmates from public facilities were matched with inmates from private facilities on the basis of classification (minimum, medium), offense category, race, prior incarcerations (0, 1, 2, or more), and age (25 or less, age 26 to 30, age 31 to 35, age 36 to 40, and age 41 or older). Of 300 releases from private prisons, only 196 matched pairs could be found. In fact, the researchers had to relax their categories to find even 196 matched pairs. The researchers also identified whether an inmate had participated in educational, vocational, substance abuse, behavioral education, or pre-release training. A seriousness of recidivism score was constructed based on the nature of the recidivism. The score ranged from 0 to 5. Zero indicated no recidivism, 1 indicated a technical violation, 2, a misdemeanor, 3, a drug or weapon possession offense, 4, a property offense, and 5, a violent or personal offense.

Lanza-Kaduce and Parker (1998) reported the following results: (1) private releasees had a lower recidivism percentage on every one of the 5 indicators except technical violations. The overall measure indicated a recidivism percentage of 17 percent for the inmates released from private prisons and 24 percent for inmates released from the public institutions. For the overall measure, this translated into a recidivism rate of 172 per 1,000 released inmates for the private institutions and 237 for the public facilities. The recidivism scale measuring seriousness indicated that public sector releasees were more likely to commit drug/weapon possession offenses, property offenses, and violent offenses. The mean level of seriousness on this 0 to 5 point scale was 3.43 for the public sector releasees and 2.32 for the private sector releasees.

Lanza-Kaduce and Parker (1998) also reported on the relative recidivism rates among releasees from the privately-operated facilities who successfully completed one or more of the programs previously listed against those who failed, dropped out, refused to participate, or were removed. Among the successful completers, 15 percent recidivated, while 40 percent of the noncompleters recidivated. We find this final result fraught with issues regarding selection bias. One can draw no conclusions about program effectiveness when you compare dropouts with those who complete a program. Under these circumstances, it is impossible to disentangle program effects from the inmate's underlying motivation to succeed (Pelissier, Rhodes, Gaes, Camp, O'Neil, Wallace, and Saylor 1998).

A critical assessment of this study has been written by the Florida Department of Corrections, Bureau of Research and Data Analysis (1998). The analysts writing this report noted four significant problems with the Lanza-Kaduce and Parker study. The first problem focuses on the putative equivalency of the private and matched public inmate releasees. The Bureau of Research and Data Analysis (BRDA) paper notes that while Lanza-Kaduce and Parker selected inmates who were minimum or medium custody, this is not the same as using the level of custody which was apparently higher, on average, for the inmates released from the public facility. Furthermore, publicly released inmates were more likely to have longer sentences, another indicator of the seriousness of the instant offense. Publicly imprisoned inmates served, on average, a much longer time than the privately held prisoners and inmates released from a public facility were more likely to have a term of supervision. This latter difference is quite important. As the BRDA researchers point out, the differences in supervision indicate that the inmates released from public institutions were more serious offenders, who were more closely monitored, thus magnifying re-arrest or other recidivism measures. This makes

the use of technical violations inappropriate. If an offender is not under supervision, he or she cannot be technically violated.

The BRDA researchers also found that offenders in the private sample had, on average, a less serious previous record. This was based on the number of prior incarcerations. Finally, the BRDA researchers were critical of the broad age categories used by Lanza-Kaduce and Parker. As the BRDA group noted, age is the most significant determinant of recidivism and either an exact match should have been used or a multivariate analysis controlling for age and some of these other variables should have been conducted.

A second, even more serious problem in the Lanza-Kaduce and Parker analysis is that 35 percent of the inmates included in the private sample had also spent a significant amount of time in a public facility. It would be impossible to disentangle the effect of the private "dose" from the public "dose." This might suggest a research design in which one compared exclusively private, exclusively public, and mixed public-private incarceration. But even this might be meaningless unless we can have assurances that the decision to select or place inmates into these facilities is not somehow entangled with their propensity to recidivate -- the problem of selection bias at a broader level.

As a third criticism, the BRDA researchers also criticized Lanza-Kaduce and Parker for using relatively small sample sizes. However, we would argue, if all of the other problems of this study could be addressed, the sample size was probably sufficient. A fourth, and more serious error, was the way in which Lanza-Kaduce and Parker procedurally measured recidivism. According to the BRDA group which had the original data, Lanza-Kaduce and Parker evaluated the recidivism of inmates released from private facilities in the period June 1, 1996, to September 30, 1996. The period for the inmates released from public facilities was January 1, 1996, to September 30, 1996. As we noted before, recidivism was assessed from the day of release until 12 months had been reached. Because there is a lag between the occurrence of a recidivism event and the recording of that event into the automated records, the recidivism differences reported by Lanza-Kaduce and Parker may be an artifact of the recording process. The recidivism data were gathered in November 1997, thus the publicly released inmates had, at most, a 23-month release period, while the privately released inmates had, at most, an 18-month release period. Depending on how long the lag is between the event and the recording of the event, this discrepancy between public and private release periods could seriously bias the results in favor of the private facility.

While this particular study of the relative differences in recidivism among privately and publicly incarcerated inmates had serious errors, we are skeptical that any such study can circumvent the problems associated with matching inmates and precluding selection bias. Aside from these methodological problems, what are the theoretical implications of such a contrast? On what theoretical basis would we expect privately and publicly operated prisons to be different in ways that would affect recidivism? If the issues revolve around programming, then it is certainly possible to evaluate inmate programs recognizing the same methodological problems with matching and selection bias. However, the public sector can also develop and deliver programs. In fact, many programs delivered in public prisons are the result of a contractual arrangement with a private provider. Finally, even if we could develop a satisfactory design to compare the recidivism rates of publicly and privately operated prisons, we would want to know what it is about the nature of operations at either the public or private facility that reduces criminality. Knowing this, we would export that knowledge to all of our prisons.

Summary of Existing Studies

For the most part, those who have evaluated private corrections in comparison to public corrections have concluded that the private correctional facilities performed as well or better than the public institutions. However, in our assessment of these evaluations, we find that most of these studies are fundamentally flawed, and we generally agree with the 1996 GAO report that there is “little information that is widely applicable to various correctional settings (p. 11).” We think this conclusion is still warranted despite the two recent evaluations conducted in Louisiana (Archambeault and Deis 1996) and Arizona (Thomas 1997a), and the recidivism study conducted by Lanza-Kaduce and Parker (1998).

In our opinion, the strongest of the existing studies are the Tennessee evaluation and the Washington State review. The Tennessee evaluation found that Corrections Corporation of America was running a prison on par with the two new facilities operated by the Tennessee Department of Corrections. The review conducted by Washington State noted that while it would be hard to generalize the findings to Washington (and, by implication, other states), it does appear that the experiences with privatization in Tennessee, Louisiana, and Florida have been positive. Even in these studies, though, there are serious shortcomings in the analyses (Thomas et al., 1996).

The most significant problem with all of these studies is that they fail to develop a coherent model of institution performance in terms of cost and quality of operations. Such a model would include the structure of the relationship between process and outcome measures. The model would also make explicit those factors that must be controlled to make institutional comparisons meaningful (Camp, Saylor and Harer 1997; Camp, Saylor and Wright 1998; Office of Research and Evaluation 1998; Saylor 1996). In this review, we have criticized researchers either for not using statistical adjustments or for incorrectly using univariate statistics. In most cases, it seems to us that researchers are inappropriately using univariate statistics to infer population differences when, in fact, they are using population data. We also were critical of Logan for not using statistics to assess the degree to which the private and public institutions were different. We think descriptive statistics could help in clarifying the strength of the differences between institutions, but this must be done in conjunction with a model that allows us to understand the relationship between outcome and process variables while simultaneously controlling for the important substantive differences among the institutions.

In Table 3, we summarize some of the other characteristics and failures of these studies. Most of the studies do not use a variety of different measurement approaches; fail to study equivalent inmate populations or have insufficient information on the comparability of the offenders; use inappropriate or no statistics; and, use a single point in time, rather than a longitudinal assessment. Most also fail to explain the nature of private innovation, the impact of privatization on the entire system, or how innovation affects performance in terms of cost and quality of operations.

Practically all of the evaluation research literature that has been produced on privatization has been designed to compare the relative performance of the public or private operation of a prison. The issue has been framed as a competition and the scorecard has been based on cost and quality. Although this will continue to be an important issue in the future, there are other fundamental ways of framing the research questions. A second, yet complementary approach, is to ask the following two questions: How does privatization change the provision of correctional services, especially in terms of changing

operations in public sector prisons, even in public systems that are generally considered to be well run and accountable to the public interest? How does public sector management of prisons influence the operations of privately managed prisons within its jurisdiction? The ways in which these systemic changes take place depend on the approach to privatization each jurisdiction takes.

Since the majority of experience in managing inmates comes from the public sector, the private sector must begin from that knowledge base. Indeed, there has been a great deal of public sector innovation in corrections, some of it borrowed from private sector operations in other industries. Unit management, objective classification of inmates, strategic planning, and the concept of direct supervision all preceded the introduction of private sector management. There is no evidence to contradict the assertion that basic management philosophy, technology, and correctional expertise have been transferred to the private sector from the public sector. Often, there is a direct transfer of personnel and policies. The question, then, is how does the private sector add value to prison operations. One can even take a broad view of this issue and consider the increased flexibility of a system that might use privatization to alleviate crowding or handle special needs populations.

The usual response is that market pressures force the private sector into a more efficient use of resources. In particular, the market supposedly creates incentive for efficiencies in two general areas. First, labor is more efficiently utilized in the private sector. This is important because labor costs typically account for 65-70 percent of the operational costs of a prison. Second, there are efficiencies realized from more flexible purchasing practices.

However, despite the claims about cost savings and increased value, in reality there have been no empirical studies documenting innovations in the private sector in the use of labor or the purchasing of goods and services. What is needed are case studies that document the innovations developed by the private sector that produce added value in the use of labor or in purchasing practices (Camp 1998). We also need to document how the labor use and purchasing practices of public sector prisons change as a result of the dynamic interplay between public and private sectors.

It appears to us that the private sector's approach to corrections has been to build upon correctional practices that already exist in well-run public prisons. The private sector does not appear to argue that they run prisons in a dramatically different way based on different philosophies of managing inmates. However, there has been little attention given to documenting the private sector approach to innovation or to the impact of competition from the private sector on the practices of the public sector.

A corollary to these systemic questions is the issue of whether the private sector is delivering too little or the public sector is delivering too much. Does the private sector save on costs by providing fewer essential services to inmates than the public sector? If so, what are the short- and long-term consequences of this lack of services? Is there a possibility that public sector prisons provide "too much" quality to inmates? That is, do they provide services to inmates that are not as readily available to some law-abiding members of society (usually indigent citizens)? While most correctional administrators (public and private) agree that U.S. prisons should meet American Correctional Association (ACA) accreditation standards at a minimum, there is probably much less agreement as to how far above the bar set by ACA standards those prisons should operate to perform effectively. How much education is excessive? Is a community standard the appropriate standard for medical care? Should we provide inmates with job training when other poor, law-abiding civilians may have less access to such training? Privatization brings these paradoxical public policy issues into sharper focus.

Quite similar to these arguments is Harding's (1997) contention that the next generation of privatization research should look for evidence of system-wide changes. Some evidence of such changes can be found in Vagg's (1994) description of the introduction of privatization into the English penal system. Vagg, who reserves judgment on the merits of privatization, notes that the introduction of the first private remand⁴⁸ prison in England—Th&Wolds, which opened in 1992—may have improved accountability for how inmates are treated in public sector prisons as well. As he notes (Vagg 1994: 307), "... private prisons have the potential to offer improved prison regimes; and ironically, in England, they were a key factor in persuading the administration that standards were necessary, if only for the purpose of monitoring contractual compliance ...". According to Vagg, the English government had been reluctant to establish prison standards such as ACA standards which are used in the U.S. With the need to oversee the Wolds contract, performance indicators were specified that the contractor, Group 4 Remand Services, Ltd., had to meet. Standards were established for security; health, safety, and hygiene; reception, registration, and discharge; regime activities (such as the grievance procedure); inmate services; and other prison functions. The standards went so far as to specify the amount of time that prisoners should spend out of cell, a key point as there was public concern about unsentenced inmates being locked in their cells for extended periods in older remand centers under public control (see James, Bottomley, Liebling, and Clare 1997). This explication of a system change introduced by adding privately operated beds is not typical in the evaluation literature on privatization. Similar efforts must be undertaken in the United States.

There are well over 100 adult correctional institutions currently being operated by private corrections companies. Yet we have analyses on only a handful. Even if these studies were rigorous and methodologically sound, the private institutions captured in these analyses may, or may not, represent the performance of the industry as a whole.

In the next section of this critique, we propose a model designed to address many of the problems we have noted throughout this paper.

Outline of How the Taft Evaluation Meets the Criteria of a Sound Evaluation

The proposed evaluation of the experiences of the Federal Bureau of Prisons with the private Federal prison operated by Wackenhut Corrections Corporation in Taft, California addresses many of the criticisms of existing evaluations of privatization. A complete discussion of the proposed evaluation can be found elsewhere (Office of Research and Evaluation, 1998). The discussion proceeds by addressing each of the areas covered in the summary table of existing quality evaluations.

System Impact: The system impact of the private operation of the Taft prison upon other operations in the Federal Bureau of Prisons is addressed by the modeling approach that will be used in the evaluation. At about the same time that the Taft prison was built, three other low security institutions were built by the Federal Bureau of Prisons in Yazoo City, Mississippi, Elkton, Ohio, and Forrest City, Arkansas. These three institutions are the institutions against which Taft will be compared. Taft and two of the comparison institutions also have minimum-security camps as part of the facility. In

48 A remand prison in the United Kingdom functions the same as a jail in the United States criminal justice system.

this discussion, though, we focus primarily upon the main facilities that house the low-security inmates.

In order to ensure that the institutions are compared on measures that are adjusted for features that are unrelated to institutional performance, it is necessary to develop models of the outcome measures that follow the procedures suggested by Saylor (1996) and Camp and colleagues (Camp, Saylor and Harer 1997; Camp, Saylor and Wright 1998). In essence, this means that both Taft and the three BOP comparison institutions will be evaluated against the performance of all of the other low security institutions in the BOP. As such, it will be possible to see how outcome measures at the Taft institution, the three comparison institutions, and the other low security BOP institutions change over the 5 years of the evaluation. This should provide information about how BOP operations change in response to the experience of having a private-sector institution as part of the BOP.

Innovation: As has been noted previously, innovation is touted as one of the reasons that private-sector companies can operate prisons more efficiently than the public sector without sacrificing the quality of services provided to inmates and the public. However, descriptions of how private-sector operators actually “do” corrections differently are usually missing (Camp 1998). To capture this component in the Taft evaluation, a full-time, on-site researcher has been placed at Taft. It is the role of this researcher to document the more qualitative aspects of how Wackenhut operates differently than the BOP. In addition, by examining organizational charts and work patterns, a close examination of the use of staff by Wackenhut and the BOP is planned.

Measures: The evaluation of Taft will use many different sources of information. There will be an audit/compliance component. All of the institutions, Taft as well as the BOP comparison institutions, are subject to ACA accreditation during the study period. In addition, there are plans to make use of periodic reviews conducted by the BOP. The Office of Research and Evaluation also plans to gather information through the annual survey of staff. Taft will participate in the staff survey, the Prison Social Climate Survey, as do all other BOP correctional facilities. A corresponding inmate survey of the social climate will also be administered at Taft and the comparison institutions. The inmate survey of the social climate is conducted on an “as needed” basis unlike the staff survey which has been conducted every year since 1988. Official records will comprise an integral part of the evaluation data. For example, Wackenhut uses the same disciplinary process as the Federal Bureau of Prisons and enters the disciplinary data into the same centralized database used by all BOP correctional facilities. Finally, numerous and periodic site visits to Taft and the comparison institutions are already, and will continue to be, part of the data collection effort for the evaluation.

Points in Time: As has been mentioned, the time frame for the evaluation is a 5 year time period. Therefore, the study is less likely to be influenced by atypical performance by any of the institutions over a relatively short period of time.

Equivalent Facilities: The Taft facility, as well as the three formal comparison institutions, were all built on an almost identical architectural plan at about the same time. While the wardens at all four facilities have been free to enhance the physical plant, there have not been any major renovations such as the construction of new buildings or the relocation of security fences at any of the four study facilities.

Equivalent Inmates: All of the study institutions house low-security inmates as defined by the Federal Bureau of Prisons classification system. None of the institutions have units for special needs inmates, such as units for those with severe medical or psychological problems or residential drug-treatment

programs. Each of the institutions receives inmates as proscribed by BOP policy. There is no attempt to send specific types of inmates to any of the institutions. Nonetheless, assignment of inmates to the institutions is not random. For example, Taft is located near Los Angeles. Taft receives a large number of Hispanic inmates because of the BOP policy of locating inmates near their point of release when this does not conflict with other BOP management concerns. Any discrepancy between the institutions in the characteristics of the inmates, though, will be addressed in statistical models used to generate the comparison measures.

Model Approach: Statistical models will be utilized that are appropriate for the different types of data collected as part of the Taft evaluation. For example, Saylor (1996) has demonstrated how residuals from regression models can be utilized to assess whether institutions are performing better, worse, or the same as expected on measures of inmate per capita cost and staff perceptions of crowding. Camp and colleagues (Camp, Saylor and Harer 1997; Camp, Saylor and Wright 1998) have demonstrated how hierarchical linear models can be used to assess institution influences on measures obtained from staff surveys, such as organizational commitment or evaluations of institutional operations. While the details of the models employed in the respective studies are beyond the scope of this discussion, the use of the measures derived is quite straight-forward. With all of the models used in the analyses, the end result is the ability to evaluate how institutions perform relative to other institutions, *after* controlling for factors that are known to be unrelated to management performance. As an example, consider evaluations of institutional operations. Experienced and inexperienced staff at the same institution differ in the evaluations that they provide of institutional operations. If we compare two institutions for “typical” staff perceptions of institutional operations, we would certainly want to control for any differences in the institutions in the proportions of experienced and inexperienced staff providing the evaluations.

Statistical Approach: Most of the statistics employed in the Taft evaluation will be used to statistically adjust comparison measures. For the most part, the emphasis is not upon making inferences to some population of prisons. For the most part, we are dealing with the population of low-security, Federal prisons. There will be some use of inferential statistics when the data come from samples of staff or inmates at the respective institutions.

Type of Facility: All of the institutions in the Taft evaluation house only adult offenders.

Security Level: All of the institutions are designed to house low-security inmates. For the most part, that is the type of inmate that is housed at the facilities.

Gender of Inmates: All of the inmates are male.

Concluding Remarks

Most evaluations of the respective strengths and weaknesses of public and private prisons have not relied upon strong theory to guide the evaluations. In the Taft evaluation, the theory of what constitutes a “good” prison as outlined by academics such as Charles Logan (Logan 1990) is supplemented with the collective practical knowledge of practitioners in the Federal Bureau of Prisons. Management at the Bureau have developed a performance measurement system called the Executive Staff Management Indicators (ESMI) module. ESMI is organized around six goals. Within

each goal, vital functions that support the achievement of those goals are outlined. The organization of these performance measures is based upon the collective wisdom and years of corrections experience reflected in the current and past compositions of the executive management staff at the BOP.

This theoretical approach guides the choice of measures to evaluate the performance of Taft and the three comparison BOP institutions. The data used in the evaluation will come from multiple sources. Therefore it will be possible in some instances to test for consistency in conclusions suggested by reliance upon different sources of information. The data itself will be adjusted by statistical models to control for differences between the institutions that are not related to performance. This information, in turn, will be supplemented with an examination of organizational innovations in correctional practices at Taft.

The proposed evaluation of Taft avoids some of the more obvious problems that have plagued existing evaluations of privatized prisons. The proposed evaluation does not, however, circumvent all problems. For example, the evaluation is still for only one jurisdiction, the U.S. Federal Government, for one private company, Wackenhut, and for only low- and minimum-security inmates. Likewise, all of the inmates are male. The study does, we think, point the way for collecting systematic evidence about the types of changes privatization produces in the operations of public systems, how private firms innovate in providing services to inmates, and whether public and private firms can provide comparable services to inmates given the constraints faced by the Federal Bureau of Prisons. To date, we have not generally found that this type of information is readily available in existing evaluations.